

Protokoll der GBV AG Discovery

Göttingen - VZG 10:30-15.30 Uhr

23.05.2018

Stand: 11.06.2018

Teilnehmer*innen: Martin Blenkle (SuUB Bremen, Protokoll), Anne Christensen (UB Lüneburg), Volker Conradt (BSZ), Reiner Diedrichs (VZG), Berrit Genat (Technische Informationsbibliothek - TIB), Jan Frederik Maas (SUB Hamburg), Oliver Goldschmidt (UB TU Harburg), Matthias Kaun (Staatsbibliothek zu Berlin), Tamara Pianos (ZBW Kiel), Gerald Steilen (VZG)

TOP1: Stand Metadatenimporte

Die [Liste der Metadatenimporte](#) im GBV-Verbundwiki wird aktualisiert und der Bearbeitungszustand geklärt.

TOP2: PAIA/DAIA

Ab 2018 wird DAIA/PAIA aus dem zentralen Haushalt der VZ finanziert. Jährliche Kosten für die Bibliotheken entfallen (bis auf die Einrichtungsgebühr). Die DAIA2 Schnittstelle kann über den Direktzugriff auf das Ausleihsystem noch nicht alle Ausleih-relevanten Informationen liefern und muss dazu erweitert werden. LBS4 verfügt allerdings über keine Schnittstelle zur Abfrage des Ausleihstatus, die direkt nutzbare Ausleih-Informationen für ein Endnutzerechtesystem liefert. Die Logik von DAIA ist weiterhin nicht direkt kompatibel mit den Parametern von LOAN4, so dass an dieser Stelle ein erhöhter Konfigurationsaufwand notwendig wird. Geplant ist eine erste Variante für das Quartal 04.2018. (Vorher soll aber zunächst die native SIP2-Schnittstelle fertiggestellt werden.)

Die Beluga-Core-Gruppe entwickelt derzeit parallel ein Verfahren zur Anreicherung der bestehenden DAIA-Schnittstelle. Diese Aktivitäten sollen zukünftig durch die Erweiterung der DAIA-Schnittstelle zusammengefügt werden.

Folio wird es zukünftig über die Schnittstelle Codex erlauben, Apps zu entwickeln, die sämtliche Inhalte prinzipiell zur Verfügung stellen können.

TOP3: Verbundkonferenz

Der Bericht der AG Discovery soll im Rahmen des Berichts des Fachbeirats bei der Verbundkonferenz erfolgen.

TOP4: Vorstellung CrossAsia (Herr Kaun)

Live-Präsentation zur Suche des FID CrossAsia (Staatsbibliothek zu Berlin) ca. 500.000 Zugriffe auf lizenziertes Material von 2000-3000 Nutzer*innen pro Jahr. 140 lizenzpflichtige Datenbanken werden angeboten. Ein gemeinsamer Index aller Quellen lässt sich nicht realisieren, sodass die zentrale Suchmaschine des FIDs gefördert über [Pazpar2](#) betrieben wird.

Pazpar2 erlaubt jedoch nur ein eingeschränktes Relevanz Ranking. Ein Problem ist die vollständige Berücksichtigung versch. Zeichenvarianten ostasiatischer Schriftzeichen. Für 40.000 lizenzierte Medien (7.5 Mio. Seiten) des CrossAsia Textrepositories existiert auch eine experimentelle Volltextsuche. Meta-Daten für eine PDA-basierte Bestellung erlauben die nutzergenerierte Beschaffung von Titeln.

TOP 3: Verschiedenes

- Personenfacette

Oliver Goldschmidt beschreibt, dass der K10plus-Zentral Index kein Feld aufweist, das alle Autorennamen beinhaltet. Dies führt zu Problemen bei der Anzeige von Facetten.

Zur Indexierung muss in den MARC-Daten im Feld 100 und 700 der sogenannte RelatorCode auf 'aut' gesetzt sein. Dies funktioniert für Datensätze, die nach RDA katalogisiert sind und auch im Feld 100 den

RelatorCode mit 'aut' besetzt haben. Abgesehen von 'aut' werden Relator Codes derzeit nicht ausgewertet; es gibt also kein Indexfeld für andere beteiligte Personen als Autoren.

Beim CBS Export für K10plus-Zentral soll zukünftig der Relator Code für Autoren*innen auch für die älteren, nicht nach RDA katalogisierten Daten, gesetzt werden, indem für alle Personen im PICAPlus-Feld 028A und ggf. 028C der RelatorCode automatisch ergänzt wird. Dadurch sollen alle Autoren*innen in einem Indexfeld (author_browse) indexiert werden können. Gleichzeitig sollen dabei zukünftig auch die LUY Datensätze ausgefiltert werden.

- Angabe der Provenienz von Metadaten

Anne Christensen fragt an, ob es möglich und sinnvoll wäre, die Metadatenherkunft in den Systemen anzuzeigen, um intern die Kollektionen klarer zuordnen zu können oder bestimmte Daten auch besonders bei den Nutzer*innen zu bewerben.

Gerald Steilen erläutert, dass diese Informationen schon immer in K10Plus-Zentral in standardisierter Form enthalten sind. So sind die ZDB-Sigel zu den einzelnen Quellen sowohl in den Metadaten, als auch in einem eigenen Feld der Suchmaschine verfügbar.

Beispiele:

PPN 897813847

marc21-Feld: 912 a	ZDB-94-OAB	
Solr-Feld:	collection_details	ZDB-94-OAB
	collection	GVK

Lukida: <https://find.ub.uni-rostock.de/id%7Bcolon%7D897813847>

PPN 271181281

marc21-Feld: 912 a	ZDB-1-POR	
	ZDB-1-SOJ	
Solr-Feld:	collection_details	ZDB-1-POR
		ZDB-1-SOJ
	collection	GVK

Lukida: <https://find.ub.uni-rostock.de/id%7Bcolon%7D271181281>

Das Thema Qualität von Metadaten wird die AG weiter zukünftig beschäftigen. Metadatenqualität muss definiert werden. Mögliche Ausgangspunkte für eine zukünftige Diskussion in der AG wären:

- Welche Maßnahmen sind realistisch möglich, um die Metadatenqualität für die Discoverysysteme zu verbessern?
Wie kann die Katalogisierung für die Anforderungen in Discoverysystemen sensibilisiert werden?
Welche minimalen Qualitätsanforderungen können benannt werden: technisch - formal - inhaltlich?
- Wie erreiche ich eine Katalogisierung nach [FAIR](#)-Prinzipien?
- Was benötigen Nutzer und welche Folgerungen ergeben sich für Fragen der Metadatenqualität?
- Welche Kriterien gibt es für die Verständlichkeit von Metadaten (graph. Darst.)?
- Wie wird Verfügbarkeit optimal dargestellt?
- Known Item Search vs. Thematische Suche
- Wie kombiniert man flüchtige und statische Metadaten?

- Bereitstellung der K10plus Zentral-Daten für Suchmaschinen-Indexierung

Die Auffindbarkeit von Daten aus K10plus Zentral über Suchmaschinen kann nur über die lokalen Discoverysysteme erreicht werden, u.a. durch deren Gestaltung in Form von Sitemaps, die dann von den Suchmaschinen geharvestet werden können. Ob es sinnvoll ist, eigene Kataloganzeigen durch Google indexieren zu lassen, muss gesondert beurteilt werden.

- Filterung für Datensätze (z.B. Swets-Daten) unter Solr7

Nach Umstellung auf Solr7 wurden an der TU Hamburg bei externen Recherchen keine Ergebnisse mehr gefunden, weil die Syntax des Filters für Datensätze, deren Anzeige beschränkt werden muss, sich geändert hat. In diesem Zusammenhang kam die Frage auf, welche Filterung von Metadaten aus den Nationallizenzen vorgenommen werden muss.

Bei der Nutzung des K10plus-Index fragt die Verbundzentrale ab, ob der Zugriff auf Datensätze vorgefiltert werden soll oder nicht. Bei der TU Hamburg besteht ein ungefilterter Zugriff auf alle Datensätze. Das bedeutet, dass die Filterung auf zugriffsgeschützte Inhalte auf Applikationsebene erfolgen muss. Kritisch sind dabei Datensätze, die in den collection_details mit der Kennzeichnung ZDB-1-PIO versehen sind, da diese Datensätze lizenzrechtlich geschützt sind und bei einer externen Recherche nicht angezeigt werden dürfen, was im Fall der TU Hamburg wie beschrieben das Discoverysystem selbst übernehmen muss. Problematisch bleibt die Anzeige des Treffers der PIO-Datenbank an sich (<https://beluga.sub.uni-hamburg.de/vufind/Search/Results?lookfor=id%3A302471707>), da dieser Datensatz im Index mit der Kennzeichnung ZDB-1-PIO versehen ist. Im Discoverysystem der TU Hamburg wird der Treffer bei externen Recherchen folgerichtig nicht angezeigt, obwohl die Anzeige dieses Treffers grundsätzlich erlaubt wäre. In vorgefilterten Systemen besteht Zugriff auf den Datensatz.

- Normdaten

Die Nutzung von GND-Normdaten kann durch Auswertung von Normdatenrelationen zur Verbesserung des Recalls von Suchen führen. Eine Indexierung von Normdatenrelationen wie diese in bibliothekarischen OPAC-Systemen üblich ist, ist in SOLR-basierten Systemen nicht verfügbar. Ein Lösungsansatz für solche Discoverysysteme ist die Integration der Normdaten in die Lucene Synonymtabelle. Das Problem ist allerdings bisher weitgehend ungelöst.

Ansätze der Suchterweiterung durch extrahierte Synonymencluster u.a. aus der GND gibt es im BSZ und in der TIB. Zu diesem Thema werden Volker Conradt und Berrit Genat in der nächsten Sitzung berichten.

Nächster Termin: 17.01.2019